Optimal Operation of Renewable Energy Plus Storage Systems Using Reinforcement Learning

Scott F. Underwood
Department of Mechanical Engineering
Stanford University
Stanford, CA
sunder24@stanford.edu

***Abstract*—Renewable energy systems are increasingly being added to the electric grid. With these renewable energy systems come variability in power production – in order to counter this, many renewable energy systems are combined with grid-scale energy storage. The operation of these renewable plus storage systems can be optimized to maximize financial gain, as locational marginal price of electricity varies drastically with time. This paper provides a framework for determining optimal operation renewable plus storage system, using a wind plus storage system in Texas as an example. The optimal policy for a given scenario is developed through Q-learning, drawing off historical data from the Electric Reliability Council of Texas. Review of the optimal policies applied to the historical data show that the addition of a storage system to an existing wind power system results in greater revenue than operation of the wind power system on its own. In addition to the demonstrated financial benefits of renewable plus storage systems, the storage provides flexibility to the grid in the form of dispatchable energy to counteract the variability of renewable energy production.**

I. Introduction

In an effort to combat climate change, there is increased implementation of renewable energy resources on the electric grid, which in 2020 made up 6 percent and 3 percent of global electricity generation, respectively [1]. This trend is expected to continue in the United States, with falling costs, tax credits, and executive orders aiming at 100 percent carbon pollution-free electricity by 2030 [2]. Wind and solar energy pose unique challenges for power planning, primarily due to their temporal variability [3].

To combat this, there is an influx of energy storage systems on the grid, such as grid-scale batteries [4]. The primary purpose of the energy storage systems is to facilitate the implementation of renewable energy by providing dispatchable electricity to counteract the variability of renewables. However, the energy storage systems can also be used to participate in energy arbitrage, the act of buying electricity at low prices and selling at higher prices for a profit [5]. The price at which electricity is bought and sold in real-time at the grid-scale is called the locational marginal price (LMP).

To facilitate interconnection and maximize profits, many renewable energy systems are being cited alongside storage systems, as hybrid renewable plus storage systems [6]. The optimization of these systems is complex due to the variable electricity prices, uncertain power generation of the renewable energy system, and multiple charging levels of the energy storage system. It is therefore of interest to develop an algorithm to determine the optimal action in any scenario, maximizing profitability of the renewable plus storage system.

II. Methodology

The Python source code for the project is contained in a public GitHub repository[1].

A. System Design

This paper will analyze a wind plus storage (WPS) system in Texas, which is under the jurisdiction of the Electric Reliability Council of Texas (ERCOT)[2]. This region is of particular interest for WPS systems because of the excellent wind resources in Texas, leading to an influx of wind and, along with it, storage, on the grid. Specifically, this paper chooses to site the system in the West load zone as seen in Figure 1, as this aligns with the strongest wind resources in Texas [7].

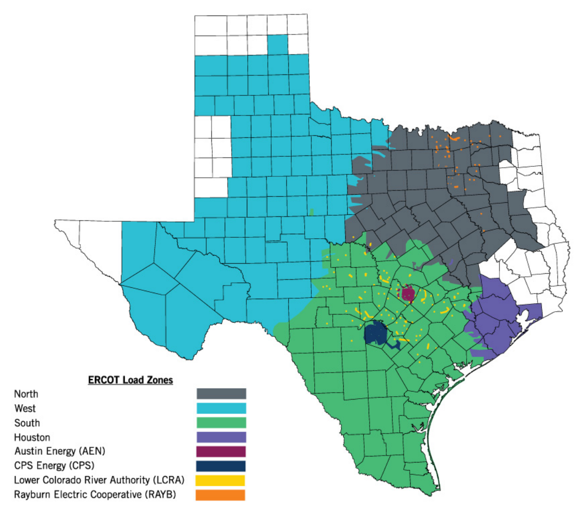

Fig. 1. ERCOT load zones[3]

[1] https://github.com/underwood-scott/CS-238-Final-Project
[2] https://www.ercot.com/
[3] https://grid-analytics.ece.utexas.edu/chart/electricity-prices

The WPS system to be analyzed will have 100 MW of wind power and 100 MW of storage, which align with typical system sizes according to the U.S. Energy Information Administration[4]. This system has two-way connection between the energy storage and the grid. There is one-way connection from the wind turbines to the grid and the energy storage. The system setup is shown in Figure 2.

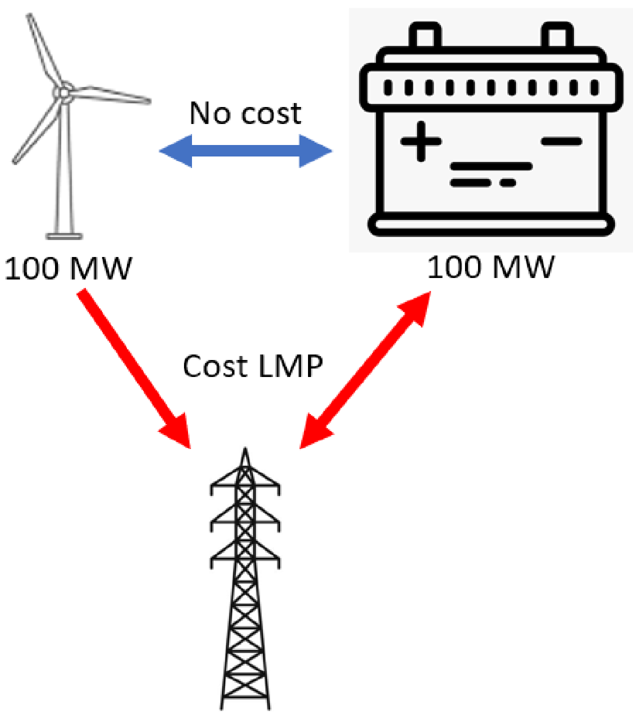

Fig. 2. WPS system setup.

To analyze this system, public ERCOT data for calendar year 2022 is used, provided in 15-minute intervals. The ERCOT data includes LMP[5] and wind power generation data[6] for the entire service region. The LMP data is filtered to the west load zone (HB_WEST). To obtain the wind power generation for the system, the total ERCOT wind generation was divided by the total ERCOT installed wind capacity of 37,000 MW[7] to get the capacity factor, and then multiplied by the capacity of 100 MW to get system generation.

B. Problem Setup

In order to determine the optimal operation of the WPS system, a modified Q-learning method is implemented, which requires the discrete state and action spaces, as well as a reward function. A model-free method such as Q-learning is optimal for this problem due to the stochaistic nature of the state space, particularly the LMP and wind power generation.

[4]https://www.eia.gov/electricity/data/eia860/
[5]https://www.ercot.com/mp/data-products/data-product-details?id=NP6-785-ER
[6]https://www.ercot.com/gridinfo/generation
[7]https://www.ercot.com/files/docs/2022/02/08/ERCOT_Fact_Sheet.pdf

1) State Space: The state space of the system is derived from three variables, represented as a NumPy array in $\mathbf{R}^3$:

1) LMP (\$/MWh): The LMP is discretized into bins, including one bin from $-\infty$ to -30, bins evenly spaced by 5 from -30 to 300, and one bin from 300 to ∞.
2) Power (MW): The power generated by the wind turbine is discretized into bins evenly spaced by 1 from 0 to 73, the maximum power generation on the year.
3) Charge (MWh): The charge of the storage is discretized into bins evenly spaced by 10 from 0 to 100.

2) Action Space: There are three possible actions for the system. To simplify calculation of the next state and the reward, any charge or discharge of the storage changes the storage charge by one bin (10 MW):

1) Discharge: Discharge the storage to the grid and sell wind power, reducing storage charge by 10 MW. This action is not allowed if the storage is in the lowest charge state. Revenue is $LMP * (P_{Wind}/4 + 10)$.
2) Hold: Hold the storage charge constant and sell the wind power. Revenue is $LMP * P_{Wind}/4$.
3) Charge: Charge the storage from the wind power, increasing the storage charge by 10 MW. Any wind power in excess of 10 MW is sold to the grid. This action is not allowed if the storage is in the highest charge state. Revenue is $LMP * (P_{Wind}/4 - 10)$

TABLE I
VARIABLE DEFINITIONS FOR WPS SYSTEM.

Input Variable	Definition	Units
Δ_{LMP}	Difference in LMP	\$/MWh
LMP	Current LMP	\$/MWh
LMP_{AVG}	Average LMP	\$/MWh
Δ_E	Change in energy	MWh
Δ_{Charge}	Change in storage charge	MWh
P_{Wind}	Wind power	MW
R	Reward from action	\$
ϵ	Exploration constant	-
$Q(s,a)$	Value function at current state-action pair	\$
$Q(s',a')$	Value function at next state-action pair	\$
γ	Discount factor	-
α	Learning rate	-

3) Reward: The reward for any state action pair is determined as the difference between the revenue from the given state/action pair and the revenue from taking the action at the average LMP. The first step to calculate the reward is to calculate the difference between the current LMP and the average LMP:

$$\Delta_{LMP} = LMP - LMP_{AVG}$$

. Next, the total change in energy into/out of the system is calculated, taking into account the wind power generated and the action taken:

$$\Delta_E = \Delta_{Charge} - \frac{P_{Wind}}{4}$$

where Δ_{Charge} is -10 MWh for action 1 above, 0 MWh for action 2, and 10 MWh for action 3. The P_{Wind} term is divided by four because the data is in 15-minute intervals, and is negative because wind power can only flow from the system to the grid.

The reward for a given state and action is then given by:

$$R = \Delta_{LMP} * \Delta_E$$

C. Q-learning Algorithm Implementation

Because there is not historical energy storage charge data for this region, the storage charge level is initialized to the middle charge level. The policy at each state is initialized to action 1 above for the highest third of LMP prices, action 2 for the middle third of LMP prices, and action 3 for the lowest third of LMP prices. The policy for the minimum charge state is initialized to action 3, and the policy for the maximum charge state is initialized to action 1. Lastly, the value functions for action 1 in the minimum charge state and action 3 in the maximum charge state are set to $-\infty$.

The algorithm iterates through each data point and chooses an action using an ϵ-greedy exploration policy, where a random action is chosen with probability ϵ, otherwise the optimal action for the state is selected. The value $\epsilon = 0.3$ was chosen to balance exploration and exploitation, taking into account the relatively small amount of data compared to the size of the state space. The value function is then updated for the given state-action pair using the following equation[8]:

$$Q(s, a) \leftarrow Q(s, a) + \alpha(R + \gamma \max_{a'} Q(s', a') - Q(s, a))$$

where $\alpha = 0.5$ in order to achieve more rapid learning, and $\gamma = 0.99$ due to the fact that the time-step between data points is 15 minutes, and there is minimal discount of monetary value in that short of a time span. To converge to an optimal policy, the algorithm iterates through the Q-learning algorithm 50 times, at which point the optimal policy for each state present in the 2022 data set using the current Q-learning approach has converged.

The algorithm then iterates one last time through the 2022 data set, implementing the optimal policy at each state and updating the charge state of the storage to reflect the action taken before moving to the next data point.

III. Results

To assess the value of adding energy storage to the wind power system, it is useful to provide a baseline case that is a wind power system without storage. This can then be compared to the WPS system to assess the added value of the energy storage.

One thing it is interesting to look at is the revenue at each time step for the two scenarios. The revenue for the scenario without storage corresponds to always selling the wind energy if the LMP is positive, and curtailing the wind if LMP is negative. This is shown below in Figure 3 for an example time period.

As seen above, the WPS system does not outperform the system without storage for all time steps. The majority of the

[8]https://algorithmsbook.com/files/dm.pdf

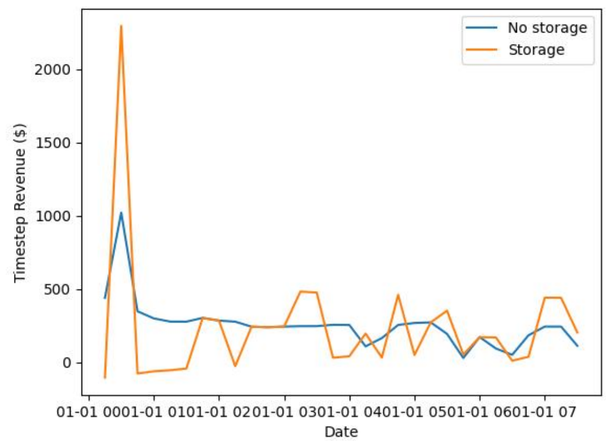

Fig. 3. Revenue per time step for January 1, 2022 from 12 am to 7 am for two power system scenarios.

benefit of the WPS system comes from exploiting periods of abnormal pricing.

Therefore, it is also important to look at the cumulative revenue throughout time for the two scenarios. The revenue for the WPS system can be calculated using the following equation:

$$Revenue = \Delta_E * LMP$$

Plotting the cumulative revenue over the year 2022 for the two scenarios in Figure 4 shows that adding the storage results in \$778,088.41 more revenue over the course of the year.

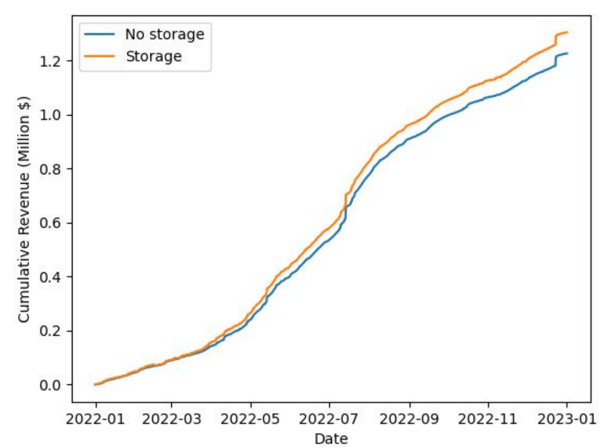

Fig. 4. Cumulative revenue of the two power system scenarios.

Considering a projected utility-scale energy storage cost of approximately \$30/kWh in 2023 [8], the 100 MWh system would cost \$30 million, giving a payback period of about 38 years, ignoring any operational expenses. However, some companies, such as Form Energy, are pledging long-duration grid-scale storage options with much lower cost, as low as

\$20/kWh[9]. With this cost, the capital cost would be \$2 million and the payback period would be reduced to 2.5 years.

It is also of interest to see the distribution of the actions throughout the year, to understand the operation of the storage and the potential degradation. The count of all actions throughout the year can be seen in Figure 5.

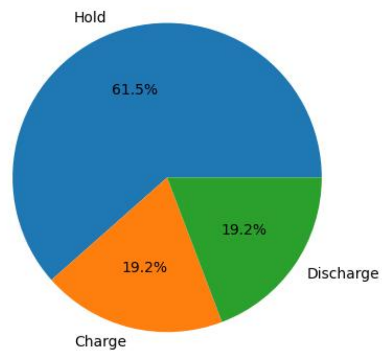

Fig. 5. Count of actions for WPS system.

The majority of the time, the policy is to hold the storage, with approximately equal occurrences of charging and discharging otherwise.

IV. Future Work

Additional work could incorporate multiple years of historical data, to provide more data points to train the model on. Computation time becomes an issue as you increase the size of the data set, but incorporating at least three years of historical data would be worthwhile to develop a more robust model.

It would also be of interest to refine the state and action space to allow for different rates of charging/discharging of the storage. While it greatly simplifies the system to only allow for constant charge/discharge rates, in reality it would result in more optimal operation to allow charging/discharging at multiple rates.

Another factor to consider for energy storage systems is storage degradation, which can greatly affect system performance after many charge/discharge cycles. It would be beneficial to implement a small penalty any time the storage is charged or discharged to account for the degradation that is taking place.

Lastly, comparing the results of Q-learning with a model-based reinforcement learning method would provide another reference point for the performance of the algorithm. Due to the easily defined reward models used, it seems that a model-based method could also work reasonably well for this particular application.

[9]https://www.cnbc.com/2021/08/25/form-energy-raises-240-million-on-iron-air-battery-promise.html

V. Conclusion

This study provides a framework for optimal operation of a WPS system. This framework can be refined and updated to provide more detail as desired, but demonstrates the financial value in adding energy storage to a renewable energy system. This model does not take into account the benefit of energy storage in providing dispatchable energy to the grid, which helps smooth the variability of renewable energy and results in even higher value for energy storage.

This framework could be applied to any variable renewable energy resource combined with storage. The framework could also be adapted to a system of any size, making it a valuable and versatile starting point in optimizing operation of renewable energy plus storage systems.

References

[1] IPCC, *Summary for Policymakers*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, pp. 3–32.
[2] "Exec. order no. 14057," dec 2021.
[3] B. N. Stram, "Key challenges to expanding renewable energy," *Energy Policy*, vol. 96, pp. 728–734, Jul. 2016. [Online]. Available: https://doi.org/10.1016/j.enpol.2016.05.034
[4] T. M. Gür, "Review of electrical energy storage technologies, materials and systems: challenges and prospects for large-scale grid storage," *Energy & Environmental Science*, vol. 11, no. 10, pp. 2696–2767, 2018.
[5] R. Walawalkar, J. Apt, and R. Mancini, "Economics of electric energy storage for energy arbitrage and regulation in new york," *Energy Policy*, vol. 35, no. 4, pp. 2558–2568, 2007. [Online]. Available: https://doi.org/10.1016/j.enpol.2006.09.005
[6] C. Budischak, D. Sewell, H. Thomson, L. Mach, D. E. Veron, and W. Kempton, "Cost-minimized combinations of wind power, solar power and electrochemical storage, powering the grid up to 99.9% of the time," *journal of power sources*, vol. 225, pp. 60–74, 2013.
[7] C. Draxl, A. Clifton, B.-M. Hodge, and J. McCaa, "The wind integration national dataset (wind) toolkit," *Applied Energy*, vol. 151, pp. 355–366, 2015. [Online]. Available: https://doi.org/10.1016/j.apenergy.2015.03.121
[8] W. Cole, A. W. Frazier, and C. Augustine, "Cost projections for utility-scale battery storage: 2021 update," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2021.